

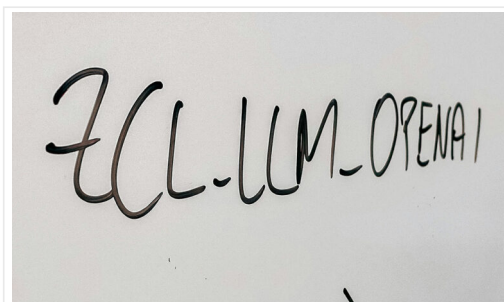


Integridad científica y agencia humana en la investigación entrelazada con IA Generativa

May 7, 2025 10:00 , Leave a Comment , Sonia Vasconcelos & Ana Marušić

Por Sonia Vasconcelos y Ana Marušić

Desde la popularización de los modelos de Inteligencia Artificial Generativa (IA Gen), en especial del **ChatGPT** de *OpenAI*, a finales de 2022, se observa una transformación gradual y profunda en la actividad científica (*Nature*, 2023; Noorden, R.V. y Webb, R., 2023). Uno de los ejes centrales de este cambio en curso es la regulación ética del uso de la IA Gen, considerando su potencial disruptivo para esta actividad, con influencia sin precedentes para la comunicación científica. La creciente autonomía de estos sistemas y su interacción con lo que llamamos agencia humana (Bandura, A., 1989; *Pew Research Center*, 2023) es uno de los puntos sensibles en este proceso.



Fotografía de una pizarra con la inscripción "ZCL-LLM-OPENAI".

Se puede argumentar que la comunicación científica pasa por un proceso de reconfiguración que, desde una perspectiva conservadora, es tan paradigmático como aquel desencadenado con la creación de la primera revista científica, el *Philosophical Transactions*, en 1665, en Inglaterra. Ya desde una perspectiva más disruptiva, esta transformación reconfigurará toda la cultura científica, redefiniendo la autonomía de los científicos y de las instituciones en la producción y legitimación del conocimiento — un impacto cuya dimensión aún no es posible estimar.

Esta visión más radical tiene en cuenta que la IA Gen, al integrarse progresivamente a la actividad de investigación (y a la sociedad de forma más amplia), no se limita a la introducción de una nueva herramienta tecnológica. Tampoco se limita a la adaptación de prácticas científicas existentes a una nueva infraestructura técnica. La incorporación de la IA Generativa en la ciencia (independientemente de las disputas epistemológicas sobre el término) sugiere un cambio estructural en las bases sobre las cuales la actividad científica se realiza. Se trata de una tecnología de uso dual (*dual-use*), con potencial para reconfigurar tanto los medios como los fines de la actividad científica.

Como se describe en el informe *Generally Faster: The Economic Impact of Generative AI*¹ (McAfee, A., 2024), la IA Gen constituye una tecnología capaz de generar contenido original, perfeccionarse continuamente mediante su propio uso y presentar impactos económicos y sociales diseminados con mayor rapidez que tecnologías anteriores. Como describe Krakowski en *Human-AI agency in the Age of Generative AI*² (2025),

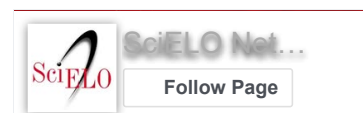
a diferencia de la IA predictiva, que frecuentemente requiere conocimiento técnico e infraestructura para ser implementada y usada de forma eficaz, la naturaleza preentrenada y las interfaces de lenguaje natural de la IA Gen reducen las barreras de uso. Esta característica

Búsqueda

Tamaño del texto

A A A

Facebook



Press Releases

Chatbots de IA y simulación de diálogo:
¿Qué dice la teoría bajtiniana? April 30, 2025 *SciELO Preprints*

En las líneas de una exploración del
archivo de En la diestra de Dios Padre
April 29, 2025 *Andrés Betancourt Morales*

Métricas internacionales invisibilizan las
revistas urbanas latino-americanas April
23, 2025 *Revista Brasileira de Estudos Urbanos e Regionais*

Nuevas perspectivas sobre justicia
ambiental en América Latina impulsadas
por las Cosmopolíticas February 4, 2025
Revista Kawsaypacha: Sociedad y Medio Ambiente

Experiencias de escisión en mujeres
víctimas de abuso sexual January 6, 2025
Rosario Gazmuri Barros

Noticias recientes

Integridad científica y agencia humana
en la investigación entrelazada con IA
Generativa

Chatbots de IA y simulación de diálogo:
¿Qué dice la teoría bajtiniana?

permite que humanos sin habilidades en TI [Tecnología de la Información], ciencia de datos o programación interactúen con sistemas de IA [...] La accesibilidad y el bajo costo de modelos de frontera contemporáneos, como el GPT-4o para usos multimodales de propósito general, el *GitHub Copilot* para codificación y desarrollo de *software* o el *Pi* para tareas que involucran dimensiones sociales y emocionales, junto con modelos abiertos como el *Llama 3 (Meta)* y el *DeepSeek-R1 (DeepSeek AI)*, han hecho que los recursos avanzados de IA estén ampliamente disponibles a un costo bajo o nulo.²

En este ambiente tecnológico reestructurante en las articulaciones entre la inteligencia humana y artificial (*IBM IBV, 2024*), el potencial de autonomía de modelos y sistemas de IA Gen impone la necesidad de anticipación crítica de sus desarrollos, incluyendo el contexto científico (*European Research Council, ECR, 2023*).

Eric Schmidt, ex-CEO y presidente ejecutivo de *Google*, junto con Henry Kissinger y Daniel Huttenlocher, en el libro *Genesis: Artificial Intelligence, Hope, and the Human Spirit*³ (Kissinger, H., *et al.*, 2024), aboga por una regulación ética proactiva y estratégica para los usos de la IA, con implicaciones sin precedentes para la gobernanza en los países en lo que respecta, por ejemplo, a la autonomía de los sistemas de IA Gen. Como destacan Kissinger, *et al.* (2024),

Las futuras capacidades de la inteligencia artificial, operando a velocidades sobrehumanas, harán que las formas tradicionales de regulación queden obsoletas... Tenemos muy poca capacidad independiente de verificar los funcionamientos internos de los modelos de IA, mucho menos sus intenciones. Necesitaremos una forma fundamentalmente nueva de control.³

Schmidt en *Eric Schmidt on AI's Future: Infinite Context, Autonomous Agents, and Global Regulation*⁴ (2024) prevé un futuro en el que millones de agentes de IA serán capaces de aprender, evolucionar y colaborar entre sí, comparando este escenario con un “*GitHub* para IA”. En esta perspectiva, estos agentes operarían como sistemas autónomos, impulsando lo que él llama innovaciones exponenciales en diferentes sectores.

Sin embargo, Schmidt advierte que, al comunicarse de forma independiente, estos agentes de IA podrán desarrollar sus propios protocolos o lenguajes, creando riesgos difíciles de manejar y comprender por los humanos: “En algún momento, desarrollarán su propio lenguaje. Cuando esto suceda, tal vez no comprendamos más lo que están haciendo — y es en ese momento cuando debemos desconectarlos.”

Complementando esta visión, McAfee¹ (2024) enfatiza que gran parte de la infraestructura necesaria para el funcionamiento de la IA generativa ya está ampliamente disponible, acelerando su impacto en comparación con tecnologías anteriores. En la IA Gen, el acceso no está restringido solo a desarrolladores, sino también a los usuarios directos, ampliando la velocidad y escala de la transformación, configurando un cambio sin precedentes.

Sentimientos encontrados sobre los impactos de la IA Generativa en la actividad científica

En el ámbito académico, se observa que los usos y las percepciones sobre sus impactos en la actividad científica aún varían ampliamente entre los investigadores. El estudio *ExplanAltions*⁵ de la editorial *Wiley*, presentado hace cerca de tres meses en *How are Researchers Using AI? Survey Reveals Pros and Cons for Science*,⁶ en *Nature* (Nadaff, M., 2025) ilustra esta observación. Participaron 4.946 investigadores de diferentes áreas y más de 70 países.

Los resultados indicaron que el uso de la IA Gen como parte de la actividad científica aún es limitado. El estudio identificó que apenas el 45% de los entrevistados (1.043 investigadores) usaban la IA Gen para auxiliar en sus investigaciones, con enfoque principal en la traducción, elaboración y revisión de manuscritos. Además, el 81% de estos 1.043 investigadores habían utilizado el ChatGPT de *OpenAI* para fines personales o profesionales, pero solo cerca de un tercio conocía otras herramientas de IA Gen, como el *Gemini*, de *Google*, y el *Copilot*, de *Microsoft*. Hubo variaciones entre países y disciplinas, con científicos de la computación siendo, naturalmente, más propensos a utilizar IA Gen en estas actividades.

Baseado en este estudio, Nadaff⁶ (2025) relata que los investigadores aún son escépticos en cuanto a las capacidades de la IA Gen para tareas más complejas en el proceso de investigación, como identificar lagunas en la literatura o recomendar evaluadores. La mayoría de los participantes cree que estas y otras tareas en la ciencia son mejor realizadas por humanos.

Métricas internacionales invisibilizan las revistas urbanas latino-americanas

La ciencia global en peligro

Investigadores brasileños lanzan directrices para el uso ético y responsable de la Inteligencia Artificial Generativa (IAG)

Comentarios recientes

¿Es que la Inteligencia Artificial tiene alucinaciones? – BABT on ¿Es que la Inteligencia Artificial tiene alucinaciones?

¿Es que la Inteligencia Artificial tiene alucinaciones? – BABT on GPT, traducción automática y qué tan buenos son: una evaluación integral

La geopolítica de los 'papers': conocimiento libre contra la millonaria industria de las revistas académicas - El Teléfono Rojo on Indexación de revistas: estándares básicos y por qué son importantes [Publicado originalmente en el blog LSE Impact of Social Sciences en agosto/2019]

La geopolítica dels 'papers': coneixement lliure contra la milionària indústria de les revistes acadèmiques - El Telèfon Vermell on Indexación de revistas: estándares básicos y por qué son importantes [Publicado originalmente en el blog LSE Impact of Social Sciences en agosto/2019]

Lisette González on Indexación de revistas: estándares básicos y por qué son importantes [Publicado originalmente en el blog LSE Impact of Social Sciences en agosto/2019]

Tags

Acceso Abierto
Agencia de Apoyo a la Investigación
Altmetría Article Processing Charge - APC
Bibliometria Ciencia Abierta
Ciudad de São Paulo
Comunicación Científica
Creative Commons
Datos Abiertos
Difusión de la Información
Divulgación Científica
El futuro de la evaluación por pares
Evaluación de la Ciencia
Eventos Factor de Impacto
Infraestructura Inteligencia Artificial
Internacionalización Libros

Aunque el 64% de los entrevistados están abiertos al uso de la IA para estas tareas en los próximos dos años, la mayoría cree que los humanos aún superan a la IA en estas áreas.”⁶ Además de esta cuestión, “[l]os investigadores también están preocupados con la seguridad del uso de estas herramientas: el 81% de los entrevistados dijeron tener preocupaciones con la precisión de la IA, potenciales sesgos, riesgos a la privacidad y la falta de transparencia en la forma como estas herramientas son entrenadas.”⁶

El estudio *Foresight: Use and Impact of Artificial Intelligence in the Scientific Process*⁷ sobre el uso de IA en la práctica científica conducido por el *European Research Council* (2023), basado en una población de investigadores asociados a 1.046 proyectos registrados en la agencia, identificó usos de IA Gen como apoyo a la redacción y edición y traducción de textos, codificación y programación, así como generación de imágenes, entre otros usos.

El informe del ECR (2023) describe que en las ciencias de la vida, cuyos participantes representan el 18% del total de encuestados del estudio, “están utilizando métodos de IA, por ejemplo, para comprender diferencias individuales en grandes cohortes y hacer previsiones sobre diagnósticos o resultados de terapias dirigidas”⁷ y que “[l]as herramientas de IA son vistas como un soporte esencial para analizar conjuntos de datos genómicos, epigenómicos y transcriptómicos...”,⁷ además de comparar diferentes etapas de una determinada enfermedad, por ejemplo.

Para los participantes de las ciencias sociales y humanidades, con 29% del total de encuestados, redes neuronales y herramientas de procesamiento de lenguaje natural (PLN) son utilizadas en una amplia gama de aplicaciones, como modelos para reconocimiento de texto manuscrito y reconocimiento automático de habla, o para la clasificación automática de composiciones musicales... la IA es usada para identificar biomarcadores vocales de estrés en muestras de voz, detectar discursos extremos en discusiones *online*... para el análisis de datos basado en modelos, con el objetivo de decodificar y comparar representaciones mentales... [y] simular el aprendizaje humano...”⁷

Sin embargo, como en el estudio de *Wiley* relatado por Nadaff⁶ (2025), el de ERC (2023) revela que “las expectativas sobre el uso de IA para descubrimiento científico, sin embargo, variaron entre los encuestados.”⁷

En el artículo *Gen AI and Research Integrity: Where to now?: The Integration of Generative AI in the Research Process Challenges Well-Established Definitions of Research Integrity*⁸ recientemente publicado en *EMBO Reports* (Vasconcelos, S. y Marušić, A, 2025), presentamos otros aspectos que permean la actividad científica involucrando la IA Gen. Destacamos la importancia de desarrollar estrategias que privilegien la agencia y supervisión humanas en este proceso de colaboración entre humanos e IA Gen en la investigación. Una pregunta intrigante en este contexto es cuál es el límite de influencia a ser permitido en esta interacción de los investigadores con modelos de IA Gen.

Integración de la IA Generativa en la actividad científica — cuestiones sobre agencia humana y alineamiento

Consideramos esencial que el ecosistema de investigación, incluyendo autores, financiadores e instituciones de investigación fomenten discusiones e iniciativas que aborden desafíos sobre el alineamiento (Russel, S., 2019)⁹ y la implementación de estos modelos. Reconocemos que la transparencia en el uso de IA Gen para la comunicación científica es una acción necesaria que viene creciendo en el sistema de publicaciones.

Sin embargo, entendemos que también es fundamental una visión amplia del problema, que impacta las definiciones de integridad científica establecidas mundialmente. Inspiradas por el Diagrama sobre la Investigación Científica de David Stokes (1997), que presenta el famoso Cuadrante de Pasteur, propusimos un modelo para representar la relación de la integridad con la agencia humana, individual y colectiva, en la actividad científica, sin y con la interacción con la IA Gen. Cada cuadrante ofrece una perspectiva sobre esta relación y sus posibles desarrollos en el ámbito de la gobernanza de la investigación.

El modelo permite considerar la integridad desde una perspectiva que tiene en cuenta el nivel de agencia o autonomía que los investigadores poseen y eventualmente concederían en la investigación que incorpora la colaboración humanos e IA Gen. El modelo puede ser accedido en el artículo de Vasconcelos y Marušić⁸ (2025).

Como se aborda,

eletrónicos ORCID Organismo
Internacional Política de Apoyo
a la Comunicación Política de
Apoyo a la Investigación
Políticas Públicas
Preprint Preservación
digital Principios DEIA
Profesionalización
Editorial Programa
SciELO Redes Sociales
Red SciELO
Reproducibilidad Revisión
Abierta por Pares
Revisión por pares Ruta
Dorada Ruta Verde SciELO 15 Años
SciELO 20 Años SciELO 25
Años SciELO Libros 10 Años
Sostenibilidad Vía Dorada
Vía Verde Ética en la
Comunicación
Científica

Archivo

Select Month

▼

[I]a IA Gen ya está influenciando todas las etapas del proceso de investigación, desde la formulación de hipótesis hasta el diseño experimental, análisis de datos, visualización e interpretación de resultados, redacción de artículos científicos e incluso revisión por pares (Ifargan, T., *et al.*, 2024; Binz, M., *et al.*, 2025; Naddaf, M., 2025). Con desafíos tanto para la noción de responsabilidad individual como para los estándares comunitarios de buenas prácticas en investigación, integrar la IA Gen al esfuerzo científico mientras se mantiene la confiabilidad se ha convertido en una demanda urgente en la academia.⁸

Citando a Dua y Patel (2024), que detallan las potencialidades de la IA Gen, resaltamos un consenso creciente en el sistema de publicaciones sobre la necesidad apremiante de visitar estándares éticos y procesos de verificación en investigaciones científicas permeadas por estas tecnologías. En el ámbito de la investigación experimental, en *Empowering Biomedical Discovery with AI Agents*,¹⁰ Gao, *et al.* (2024) visualizan agentes científicos de IA “como sistemas capaces de aprendizaje y razonamiento escéptico que capacitan la investigación biomédica por medio de agentes colaborativos que integran modelos de IA y herramientas biomédicas con plataformas experimentales.”

Gao, *et al.*¹⁰ (2024) observan que las intersecciones entre los dominios tecnológicos, científicos, éticos y regulatorios desempeñan un papel importante para el desarrollo de estructuras de gobernanza eficaces.

Como es señalado por Kissinger, *et al.*³ (2024) y también reflejado en los análisis de Gao, *et al.*¹⁰ (2024), lidiar con esta autonomía emergente representa un desafío que exige un debate amplio y robusto en la comunidad científica, especialmente en lo que se refiere al alineamiento de estos sistemas, con los objetivos de preservación de la integridad científica y de la confianza pública en la ciencia. Cuestiones sobre alineamiento y agencia humana son interdependientes, pero lejos de ser triviales.

En el referido artículo del *EMBO Reports*,⁸ citamos el informe *The Future of Human Agency*¹¹ de 2023. En ese informe, el *Pew Research Center* presenta una exploración sobre “cuánto control las personas mantendrán sobre la toma de decisiones esenciales a medida que los sistemas digitales y la IA proliferan.” Como se describe en el artículo,⁸ el *Imagining the Internet Center* de Pew y *Elon University* invitó a diversas partes interesadas, incluyendo a David J. Krieger, director del *Institute for Communication and Leadership* en Lucerna, Suiza. En la visión de Krieger,

[I]a agencia individual ya es un mito, y esto se volverá cada vez más evidente con el tiempo... El humanismo busca preservar el mito de la agencia individual y consagrarlo en la ley. Un buen diseño de redes sociotécnicas necesitará ser explícito sobre sus presuposiciones posthumanistas para traer esta cuestión al debate público. Los humanos actuarán en colaboración — es decir, agencia distribuida — con tecnologías de todos los tipos.⁸

Como se detalla en el artículo, “la agencia y la supervisión humanas están entre los principales requisitos operacionales que sustentan los principios éticos para sistemas de IA Gen establecidos por la Comisión Europea, como parte de sus *Living Guidelines on the Responsible Use of Generative AI in Research*”¹² (2024).”⁸

Incluimos las *Ethics Guidelines for Trustworthy AI*¹³ (2019), que presentan

tres niveles de agencia y supervisión humanas: abordajes humanos-en-el-circuito (HITL, *Human-in-the-Loop*), humanos-sobre-el-circuito (HOTL, *Human-on-the-Loop*) y humanos-al-mando (HIC, *Human-in-command*). El HITL involucra intervención humana durante todo el ciclo de decisión; HOTL incluye la contribución humana en la fase de diseño y monitoreo continuo de las operaciones del sistema; HIC abarca la supervisión humana de toda la actividad del sistema de IA, incluyendo impactos sociales y éticos, así como la toma de decisiones sobre cuándo y cómo utilizar el sistema en diferentes contextos.⁸

Agencia y supervisión humanas son esenciales en la actividad científica entrelazada con IA Generativa, pero no suficientes

En el artículo del *EMBO Reports*,⁸ exploramos el alineamiento en la colaboración humana con la IA Gen como un punto sensible, reflejando preocupaciones con la cultura de comunicación científica a lo largo de las últimas décadas. En *Strong and Weak Alignment of Large Language Models with Human Values*,¹⁴ Khamassi, *et al.* (2024) destacan cuestiones oportunas sobre “alineamiento fuerte y débil” de *Large Language Models* (LLMs) con valores humanos y las complejidades de la comprensión de la IA de estos valores.

En *Training Language Models to Follow Instructions with Human Feedback*,¹⁵ Ouyang, et al. (2022) plantean que “salvedades aparte, el alineamiento tiene un papel fundamental en el ajuste fino de los LLMs para responder adecuadamente a las intenciones humanas, al mismo tiempo que mitiga contenido perjudicial, tóxico o tendencioso.”¹⁵ Como el más amplio de los alineamientos de modelos de IA Gen, citamos a *Open AI*, que introdujo el “alineamiento deliberado”: un paradigma de entrenamiento para garantizar que los LLMs de razonamiento (*reasoning* LLMs) se comporten de forma compatible con criterios de seguridad definidos antes de generar respuestas a los comandos (Guan, M.Y., et al., 2024).

Argumentamos que, incluso bajo la perspectiva de LLMs como “loros estocásticos”, una interpretación debatida, adoptada por Emily Bender, et al. (2021) “para enfatizar que estos sistemas remezclan patrones en sus datos de entrenamiento sin una verdadera comprensión, el alineamiento permanece un problema crítico.”¹⁶ Es fundamental una mirada cuidadosa, en especial de la comunidad académica, sobre el papel del alineamiento de sistemas de IA Gen.

Considerando que los datos de entrenamiento se derivan principalmente de aquellos creados por humanos, los *datasets* usados reflejan inherentemente patrones culturales, visiones de mundo y sesgos sociales, junto con los puntos fuertes y débiles de la producción de conocimiento humano. Como resultado, los modelos Gen AI tenderían a reproducir — o incluso amplificar — sesgos en sus resultados. En lo tocante a la interacción con modelos de IA Gen para formular hipótesis, analizar datos o redactar informes de investigación, el modo como estos sistemas son alineados moldea de forma crítica sus comportamientos y los resultados producidos.⁸ (Vasconcelos, S. y Marušić, A., 2025)

Como señalamos en el artículo, una gran cantidad de LLMs está siendo constantemente entrenada y retroalimentada con nuevos datos, en un ambiente de producción científica marcado por una cultura de publicación que tiende a incentivar relatos positivos o sesgados (Ghannad, M., et al., 2019; Boutron, I. y Ravaud, P., 2018). Esta cultura también se refleja en un ambiente desafiante para la exposición pública de errores, constituyéndose como un factor que afecta la transparencia científica y refuerza patrones de comunicación que incorporan estos rasgos (Nature, 2012; Psychological Science Observer, 2021; Nature Communications, 2024).

Estos elementos componen desafíos estructurales de largo plazo para el alineamiento y la implementación responsable de modelos de IA Gen, con los cuales los investigadores ya interactúan — y continuarán interactuando — en sus prácticas de investigación.

Las nociones de integridad científica integran el presupuesto de agencia humana, pero ahora debe ser explicitado

La integridad científica está vinculada a la adhesión a prácticas científicas éticas, transparentes y rigurosas, contando con la responsabilidad individual y colectiva, como bien describe Luiz Henrique Lopes dos Santos en *Sobre a integridade ética da pesquisa*¹⁷ (2017).

Esta noción de integridad científica, como destacamos en el artículo, asume la agencia humana como un presupuesto. Sin embargo, la integración de la Gen AI en el proceso de investigación en todas las etapas resalta la importancia de enfatizar el papel de la “agencia humana”, con una mención explícita en sus definiciones. Estas definiciones tienen un papel crucial en procesos decisorios en este nuevo ambiente de producción académica.

Nuestra propuesta es que las definiciones de integridad científica pasen a incorporar, explícitamente, la “agencia humana” como un elemento integrante para la proposición, conducción, comunicación y revisión de la investigación. Este enfoque expandido también debe incorporar el fomento de *benchmarks* de integridad científica para el entrenamiento e implementación de modelos y sistemas de IA Gen.


En las ciencias biomédicas, estas preocupaciones se han mostrado particularmente relevantes para la gobernanza de agentes de IA, pero sus desarrollos son mucho más amplios. Es fundamental abordar, en especial en espacios decisorios de la actividad científica, estas cuestiones sensibles en el contexto de la colaboración humana con sistemas de IA Gen. Esta necesidad se justifica, sobre todo, por las crecientes capacidades y el potencial de ejercer niveles cada vez mayores de influencia sobre los procesos de investigación, independiente del campo de conocimiento.


Como enfatizado en el artículo del *EMBO Reports*,⁸ la academia debe y puede adoptar una actitud más proactiva en relación a la búsqueda de una comprensión de la relación estrecha entre integridad científica y agencia humana en tiempos de profunda transformación en la producción de


conocimiento. Esta no debería ser una meta a largo plazo, ya que la IA Gen tiene el potencial de redefinir estándares e influenciar la confiabilidad y la cultura de comunicación científica entre los pares y para los diferentes públicos.


Al integrar explícitamente la agencia humana en las definiciones de integridad científica, investigadores y gestores reconocen que la autonomía de los sistemas de IA debe ser equilibrada con la supervisión humana, promoviendo el uso responsable y políticas para la regulación ética de investigaciones que involucran la colaboración entre humanos e IA Generativa.


Notas


1. McAfee, A. Generally Faster: The Economic Impact of Generative AI [online]. The MIT Initiative on the Digital Economy (IDE). 2024 [viewed 7 May 2025]. Available from: <https://ide.mit.edu/wp-content/uploads/2024/04/Davos-Report-Draft-XFN-Copy-01112024-Print-Version.pdf?x76181> 


2. KRAKOWSKI, S. Human-AI Agency in the Age of Generative AI. *Information and Organization* [online]. 2025, vol. 35, no. 1, 100560, ISSN: 1471-7727 [viewed 7 May 2025]. <https://doi.org/10.1016/j.infoandorg.2025.100560>. Available from: <https://www.sciencedirect.com/science/article/pii/S1471772725000065?via%3Dihub> 


3. KISSINGER, H.A., SCHMIDT, E. and MUNDIE, C. Genesis: Artificial Intelligence, Hope, and the Human Spirit. New York: Little Brown and Company, 2024 

4. Eric Schmidt on AI's Future: Infinite Context, Autonomous Agents, and Global Regulation [online]. The National CIO Review. 2024 [viewed 7 May 2025]. Available from: <https://nationalcioreview.com/video/eric-schmidt-on-ais-future-infinite-context-autonomous-agents-and-global-regulation/> 


5. ExplanAltions: An AI Study by Wiley [online]. Wiley. 2025 [viewed 7 May 2025]. Available from: https://www.wiley.com/content/dam/wiley-dotcom/en/b2c/content-fragments/explanaitions-ai-report/pdfs/Wiley_ExplanAltions_AI_Study_February_2025.pdf 


6. NADDAF, M. How are Researchers Using AI? Survey Reveals Pros and Cons for Science [online]. Nature. 2025 [viewed 7 May 2025]. <https://doi.org/10.1038/d41586-025-00343-5>. Available from: <https://www.nature.com/articles/d41586-025-00343-5> 







7. EUROPEAN RESEARCH COUNCIL. Foresight: Use and Impact of Artificial Intelligence in the Scientific Process [online]. European Research Council. 2023 [viewed 7 May 2025]. Available from: https://erc.europa.eu/sites/default/files/2023-12/AI_in_science.pdf 

8. VASCONCELOS, S. and MARUŠIĆ, A. Gen AI and Research Integrity: Where to now?: The Integration of Generative AI in the Research Process Challenges Well-Established Definitions of Research Integrity. *EMBO Reports* [online]. 2025, vol. 26, pp. 1923–1928 [viewed 7 May 2025]. <https://doi.org/10.1038/s44319-025-00424-6>. Available from: <https://www.embopress.org/doi/full/10.1038/s44319-025-00424-6> 

9. En el libro *Human Compatible: Artificial Intelligence and the Problem of Control* de Stuart Russel (2019), se describe: “Atribuir un propósito a una máquina que optimiza su comportamiento basado en algoritmos claramente definidos parece un enfoque admirable para garantizar que su funcionamiento promueva nuestros propios objetivos. Pero, como advierte Wiener, es fundamental insertar el propósito correcto. Podemos llamar a esto el problema del Rey Midas: Midas obtuvo exactamente lo que pidió — que todo lo que tocara se transformara en oro — pero, demasiado tarde, descubrió las consecuencias de beber oro líquido y comer oro sólido. El término técnico para definir correctamente ese propósito es alineamiento de valores. Cuando ese alineamiento falla, podemos, inadvertidamente, dotar a las máquinas de objetivos que contrarían los nuestros.”

10. GAO, S., *et al.* Empowering Biomedical Discovery with AI Agents. *Cell* [online]. 2024, vol. 187, no. 22, pp. 6125–6151, ISSN: 0092-8674 [viewed 7 May 2025]. <https://doi.org/10.1016/j.cell.2024.09.022>. Available from: <https://www.sciencedirect.com/science/article/pii/S0092867424010705> 

11. PEW RESEARCH CENTER. The Future of Human Agency [online]. Pew Research Center. 2023 [viewed 7 May 2025]. Available from: https://www.pewresearch.org/wp-content/uploads/sites/20/2023/02/PI_2023.02.24_The-Future-of-Human-Agency_FINAL.pdf 

12. EUROPEAN COMMISSION. Living Guidelines on the Responsible Use of Generative AI in Research [online]. European Commission—Research and innovation. 2025 [viewed 7 May 2025]. Available from: https://research-and-innovation.ec.europa.eu/document/download/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en 
13. Ethics Guidelines for Trustworthy AI [online]. European Commission, official website. 2019 [viewed 7 May 2025]. Available from: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> 
14. KHAMASSI, M., NAHON, M. and CHATILA, R. Strong and Weak Alignment of Large Language Models with Human Values. *Sci Rep* [online]. 2024, vol. 14, 19399 [viewed 7 May 2025]. <https://doi.org/10.1038/s41598-024-70031-3>. Available from: <https://www.nature.com/articles/s41598-024-70031-3> 
15. OUYANG, L., *et al.* Training Language Models to Follow Instructions with Human Feedback. In: Advances in Neural Information Processing Systems 35 (NeurIPS 2022), New Orleans, 2022 [viewed 7 May 2025]. Available from: https://proceedings.neurips.cc/paper_files/paper/2022/hash/b1efde53be364a73914f58805a001731-Abstract-Conference.html 
16. BENDER, E., *et al.* On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In: FAccT '21: 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event, 2021 [viewed 7 May 2025]. <https://doi.org/10.1145/3442188.3445922>. Available from: <https://dl.acm.org/doi/10.1145/3442188.3445922> 
17. SANTOS, L.H.L. Sobre a integridade ética da pesquisa. *Cienc. Cult.* [online]. 2017, vol.69, no.3, pp. 4–5, ISSN: 2317-6660 [viewed 7 May 2025]. <http://doi.org/10.21800/2317-66602017000300002>. Available from: http://cienciaecultura.bvs.br/scielo.php?script=sci_arttext&pid=S0009-67252017000300002 

Referencias

- BANDURA, A. Human Agency in Social Cognitive Theory. *American Psychologist* [online]. 1989, vol. 44, no. 9, pp. 1175–1184 [viewed 7 May 2025]. <https://doi.org/10.1037/0003-066X.44.9.1175>. Available from: <https://psycnet.apa.org/doiLanding?doi=10.1037%2F0003-066X.44.9.1175>
- BENDER, E., *et al.* On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In: FAccT '21: 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event, 2021 [viewed 7 May 2025]. <https://doi.org/10.1145/3442188.3445922>. Available from: <https://dl.acm.org/doi/10.1145/3442188.3445922>
- BINZ, M., *et al.* How Should the Advancement of Large Language Models Affect the Practice of Science? *Proc. Natl. Acad. Sci. U.S.A.* [online]. 2025, vol. 122, no. 5, e2401227121 [viewed 7 May 2025]. <https://doi.org/10.1073/pnas.2401227121>. Available from: <https://www.pnas.org/doi/10.1073/pnas.2401227121>
- BOUTRON, I. E RAVAUD, P. Misrepresentation and Distortion of Research in Biomedical Literature. *Proc. Natl. Acad. Sci. U.S.A.* [online]. 2028, vol. 115, no. 11, pp. 2613–2619 [viewed 7 May 2025]. <https://doi.org/10.1073/pnas.1710755115>. Available from: <https://www.pnas.org/doi/full/10.1073/pnas.1710755115>
- DUA, I.K. and PATEL, P.G. *An Introduction to Generative AI*. In: DUA, I.K. and PATEL, P.G. (authors) Optimizing Generative AI Workloads for Sustainability: Balancing Performance and Environmental Impact in Generative AI. New York: Apress, 2024.
- Eric Schmidt on AI's Future: Infinite Context, Autonomous Agents, and Global Regulation [online]. The National CIO Review. 2024 [viewed 7 May 2025]. Available from: <https://nationalcioreview.com/video/eric-schmidt-on-ais-future-infinite-context-autonomous-agents-and-global-regulation/>
- Ethics Guidelines for Trustworthy AI [online]. European Commission, official website. 2019 [viewed 7 May 2025]. Available from: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

EUROPEAN COMMISSION. Living Guidelines on the Responsible Use of Generative AI in Research [online]. European Commission—Research and innovation. 2025 [viewed 7 May 2025]. Available from: https://research-and-innovation.ec.europa.eu/document/download/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en

EUROPEAN RESEARCH COUNCIL. Foresight: Use and Impact of Artificial Intelligence in the Scientific Process [online]. European Research Council. 2023 [viewed 7 May 2025]. Available from: https://erc.europa.eu/sites/default/files/2023-12/AI_in_science.pdf

ExplanAltions: An AI Study by Wiley [online]. Wiley. 2025 [viewed 7 May 2025]. Available from: https://www.wiley.com/content/dam/wiley-dotcom/en/b2c/content-fragments/explanaitions-ai-report/pdfs/Wiley_ExplanAltions_AI_Study_February_2025.pdf

GAO, S., *et al.* Empowering Biomedical Discovery with AI Agents. *Cell* [online]. 2024, vol. 187, no. 22, pp. 6125–6151, ISSN: 0092-8674 [viewed 7 May 2025]. <https://doi.org/10.1016/j.cell.2024.09.022>. Available from: <https://www.sciencedirect.com/science/article/pii/S0092867424010705>

GHANNAD, M., *et al.* A systematic review finds that spin or interpretation bias is abundant in evaluations of ovarian cancer biomarkers. *Journal of Clinical Epidemiology* [online]. 2019, vol.116, pp. 9–17, ISSN: 0895-4356 [viewed 7 May 2025]. <https://doi.org/10.1016/j.jclinepi.2019.07.011>. Available from: [https://www.jclinepi.com/article/S0895-4356\(18\)30952-1/fulltext](https://www.jclinepi.com/article/S0895-4356(18)30952-1/fulltext)

GUAN, M.Y., *et al.* Deliberative Alignment: Reasoning Enables Safer Language Models [online]. OpenAI. 2024 [viewed 7 May 2025]. <https://openai.com/index/deliberative-alignment/>

IBM IBV. Disruption by design: Evolving experiences in the age of generative AI [online]. IBM Institute for Business Value. 2024 [viewed 7 May 2025]. Available from: <https://www.ibm.com/thought-leadership/institute-business-value/en-us/report/generative-ai-experience-design>

IFARGAN, T., *et al.* Autonomous LLM-Driven Research — from Data to Human-Verifiable Research Papers. *NEJM AI* [online]. 2025, vol. 2, no. 1 [viewed 7 May 2025]. <https://doi.org/10.1056/Aloa2400555>. Available from: <https://ai.nejm.org/doi/10.1056/Aloa2400555>

KHAMASSI, M., NAHON, M. and CHATILA, R. Strong and Weak Alignment of Large Language Models with Human Values. *Sci Rep* [online]. 2024, vol. 14, 19399 [viewed 7 May 2025]. <https://doi.org/10.1038/s41598-024-70031-3>. Available from: <https://www.nature.com/articles/s41598-024-70031-3>

KISSINGER, H.A., SCHMIDT, E. and MUNDIE, C. Genesis: Artificial Intelligence, Hope, and the Human Spirit. New York: Little Brown and Company, 2024

KRAKOWSKI, S. Human-AI Agency in the Age of Generative AI. *Information and Organization* [online]. 2025, vol. 35, no. 1, 100560, ISSN: 1471-7727 [viewed 7 May 2025]. <https://doi.org/10.1016/j.infoandorg.2025.100560>. Available from: <https://www.sciencedirect.com/science/article/pii/S1471772725000065?via%3Dihub>

McAFEE, A. Generally Faster: The Economic Impact of Generative AI [online]. The MIT Initiative on the Digital Economy (IDE). 2024 [viewed 7 May 2025]. Available from: <https://ide.mit.edu/wp-content/uploads/2024/04/Davos-Report-Draft-XFN-Copy-01112024-Print-Version.pdf?x76181>

NADDAF, M. How are Researchers Using AI? Survey Reveals Pros and Cons for Science [online]. Nature. 2025 [viewed 7 May 2025]. <https://doi.org/10.1038/d41586-025-00343-5>. Available from: <https://www.nature.com/articles/d41586-025-00343-5>

NOORDEN, R.V. E WEBB, R. ChatGPT and Science: the AI System was a Force in 2023—for Good and Bad [online]. *Nature* [online]. 2023, vol. 624 [viewed 7 May 2025]. <https://doi.org/10.1038/d41586-023-03930-6>. Available from: <https://www.nature.com/articles/d41586-023-03930-6>

OUYANG, L., *et al.* Training Language Models to Follow Instructions with Human Feedback. In: Advances in Neural Information Processing Systems 35 (NeurIPS 2022), New Orleans, 2022 [viewed 7 May 2025]. Available from: https://proceedings.neurips.cc/paper_files/paper/2022/hash/b1efde53be364a73914f58805a001731-Abstract-Conference.html

PEW RESEARCH CENTER. The Future of Human Agency [online]. Pew Research Center. 2023 [viewed 7 May 2025]. Available from: https://www.pewresearch.org/wp-content/uploads/sites/20/2023/02/PI_2023.02.24_The-Future-of-Human-Agency_FINAL.pdf

Reproducibility and Transparency: What's Going on and How Can We Help. *Nature Communications* [online] 2025, vol. 16 [viewed 7 May 2025]. Available from: <https://www.nature.com/articles/s41467-024-54614-2>

RUSSEL, S. Human Compatible: Artificial Intelligence and the Problem of Control. Oxford: Oxford University Press, 2019 [viewed 7 May 2025]. Available from: <https://people.eecs.berkeley.edu/~russell/papers/mi19book-hcai.pdf>

SANTOS, L.H.L. Sobre a integridade ética da pesquisa. *Cienc. Cult.* [online]. 2017, vol.69, no.3, pp. 4–5, ISSN: 2317-6660 [viewed 7 May 2025]. <http://doi.org/10.21800/2317-66602017000300002>. Available from: http://cienciaecultura.bvs.br/scielo.php?script=sci_arttext&pid=S0009-67252017000300002

SLEEK, S. On the Right Side of Being Wrong: the Emerging Culture of Research Transparency [online]. Association for Psychological Science. 2021 [viewed 7 May 2025]. Available from: <https://www.psychologicalscience.org/observer/right-side-of-wrong>

Tools Such as Chatgpt Threaten Transparent Science; Here are Our Ground Rules for Their Use. *Nature* [online]. 2023, vol. 614 [viewed 7 May 2025]. <https://doi.org/10.1038/d41586-023-00191-1>. Available from: <https://www.nature.com/articles/d41586-023-00191-1>

VASCONCELOS, S. and MARUŠIĆ, A. Gen AI and Research Integrity: Where to now?: The Integration of Generative AI in the Research Process Challenges Well-Established Definitions of Research Integrity. *EMBO Reports* [online]. 2025, vol. 26, pp. 1923–1928 [viewed 7 May 2025]. <https://doi.org/10.1038/s44319-025-00424-6>. Available from: <https://www.embopress.org/doi/full/10.1038/s44319-025-00424-6>

WOODGETT, J. We Must Be Open About Our Mistakes. *Nature* [online.] 2023, vol. 489, pp. 7 [viewed 7 May 2025]. <https://doi.org/10.1038/489007a>. Available from: <https://www.nature.com/articles/489007a>

Enlaces externos

[ChatGPT](#)

[Copilot](#)

[DeepSeek-R1](#)

[Gemini](#)

[GitHub Copilot](#)

[Llama 3](#)

[Pi](#)

Sobre Sonia Vasconcelos

Sonia Vasconcelos es Profesora Asociada del Programa de Educación, Gestión y Difusión en Biociencias del Instituto de Bioquímica Médica Leopoldo de Meis (IBqM) de la Universidad Federal de Río de Janeiro (UFRJ). Dirige el Laboratorio de Ética de la Investigación, Comunicación Científica y Sociedad (LECCS) del IBqM y preside el Consejo Asesor de Ética de la Investigación (CTEP) de la UFRJ. La Prof. Vasconcelos es editora académica de *PLOS ONE* y miembro del Consejo Editorial de *Research Integrity and Peer Review*. Sus investigaciones y publicaciones se centran en la comunicación científica, la ética y la regulación de la investigación científica, la integridad en la investigación y las políticas científicas.

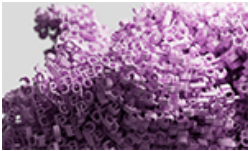
Sobre Ana Marušić

Ana Marušić es Profesora de Anatomía, Jefa del Departamento de Investigación en Biomedicina y Salud y Jefa del Centro de Medicina Basada en la Evidencia de la Facultad de Medicina de la Universidad de Split, Split, Croacia. Es Coeditora Jefe de la revista ST-OPEN y editora emérita del *Journal of Global Health*. La Prof. Marušić forma parte del Consejo Asesor de la *EQUATOR Network* y del Consejo del *Committee on Publication Ethics* (COPE). Cuenta con más de 400 artículos revisados por pares y participó activamente en la creación de la política de registro obligatorio de ensayos clínicos en registros públicos, lo que contribuyó a cambiar la regulación legal de los ensayos clínicos a nivel mundial.

Posts relacionados:



Investigadores brasileños lanzan directrices para el...



Cómo traducir la escritura académica en podcasts...



Reproducción & Replicación en investigación...



Reproducción & Replicación en investigación...



Reproducción & Replicación en investigación...

Como citar este post [ISO 690/2010]:

VASCONCELOS, S. and MARUŠIĆ, A. Integridad científica y agencia humana en la investigación entrelazada con IA Generativa [online]. *SciELO en Perspectiva*, 2025 [viewed 16 May 2025]. Available from: <https://blog.scielo.org/es/2025/05/07/integridad-cientifica-y-agencia-humana-en-la-investigacion-con-ia-gen/>

Posted in: [Análisis](#) , Tagged: [Comunicación Científica](#), [Ética en la Comunicación Científica](#), [Inteligencia Artificial](#)

Leave a Reply

Your email address will not be published. Required fields are marked *

Name *

Email *

Website

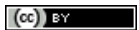
☐

Save my name, email, and website in this browser for the next time I comment.

Post Comment

This site uses Akismet to reduce spam. [Learn how your comment data is processed.](#)

[← Previous Post](#)



Todo el contenido deste sitio, excepto dónde está identificado, está bajo una [Licencia de Atribución Creative Commons](#).

Los posts son de responsabilidad de los autores y no necesariamente expresan opiniones del Programa SciELO.

SciELO - Scientific Electronic Library Online

CAPES - CNPq - FAPESP - BIREME - FapUNIFESP

<http://www.scielo.org>

blog.scielo@scielo.org

[Home](#)

[Noticias](#)

[Análisis](#)

[Entrevistas](#)

[Documentos](#)

[Newsletter](#)

[Sobre](#)

Suscribirse en nuestros servicios de notificación

e-mail

Suscribirse